DOCUMENT RESUME

ED 219 103 JC 820 346

AUTHOR Story, Sherie; Swanson, David
TITLE Getting at the Factors Underlying Trends Using Statistical Decomposition Techniques.
INSTITUTION Washington State Board for Community Coll. Education, Olympia.
PUB DATE [82]
NOTE 13p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Community Colleges; *Educational Trends; *Employment Patterns; *Enrollment Trends; Mathematical Models; Organizational Change; *Statistical Analysis; Statistical Bias; *Trend Analysis; Two Year Colleges
IDENTIFIERS *Decomposition Analysis (Statistics); *Washington

ABSTRACT

This paper presents and illustrates a technique that can be used to analyze the internal and external influences on community college enrollment trends and staffing patterns in a way that quantitatively expresses the amount of increase or decrease attributable to various competing factors. Introductory material relates the rationale behind the Washington State Board for Community Colleges' choice of a statistical approach for its comprehensive review of the status of the community college system. Next, a formal mathematical presentation of the descriptive statistical approach chosen, which is based on decomposition analysis, is presented, followed by an outline of the technique's limitations, particularly the fact that the order in which variables are introduced affects the outcome of the analysis. Then three examples, using actual data from the Washington community college system, highlight the application of the technique and its value in ascribing weights to different factors in explaining: (1) changes in community college participation rate over time and the contribution to these changes by the population's sex-structure, age-structure by sex, and participation rate structure; (2) differences in participation rates between two community college service areas; and (3) changes in faculty/student ratios at a specific college over time. The paper concludes with a discussion of how the technique can be used to analyze changes in institutional practices. (HB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219103

GETTING AT THE FACTORS UNDERLYING TRENDS USING STATISTICAL DECOMPOSITION TECHNIQUES

Sherie Story and David Swanson
Washington State Board for Community College Education
Olympia, Washington

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R. Bell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Abstract

Trends are typically expressed in terms of absolute or percentage values, e.g., "enrollment increased by 10 percent last year." Although descriptors such as these are easily calculated and interpreted, they say nothing about the component's underlying trends. This paper presents a powerful, yet easily-used technique, that offers a quantitative approach for doing this. This technique, which has not found routine application to educational data, is drawn from the areas of statistical decomposition analysis and information theory, and offers easy-to-interpret results. It can be applied at various levels within the educational environment, ranging from departments to entire college systems. The method is formally described and explained with the help of an example. Following this, data relating to community colleges in Washington State are used to illustrate how the technique quantifies the effects of internal and external influences on enrollment trends and staffing patterns. This illustrative analysis leads to a discussion of how these influences, once identified and measured, can be used to analyze changes in institutional practices.

JC 820 346

INTRODUCTION

The idea for this paper grew out of a practical need. This past year our board was engaged in a comprehensive review and analysis of the status of the community college system. A substantial amount of statistical information was compiled as a part of the study and, naturally, one item was the increased participation rate in community college programs by the citizens of the state. Of particular interest were the factors that had contributed to the increase. Although a quick examination of the "raw" figures on female enrollment, part-time enrollment, etc., provided fair general indications, we wanted to present the information in a way that quantitatively expressed the amount of the increase that was attributable to the various contributing factors. We reviewed the statistical techniques available for doing this, including those based on regression analysis, log-iinear analysis, information theory, and descriptive statistics. We chose the descriptive statistical approach--primarily because of its ease of application and also the concepts and results are relatively easy to explain.

In this paper we are going to discuss this approach to analyzing components underlying differences or changes in rates. First, the technique will be formally defined using some mathematical notation. Example applications using actual data will then be used to illustrate how this technique quantifies the effects of various factors on changes and differences as follows:

1. Changes in the overall state community college participation rate over time;

2. Differences in participation rates between two community college service areas; and,

3. Changes in faculty/student ratios at a specific college over time.

Following the third example is a more detailed discussion of the interpretation and analysis of the results obtained through application of the techniques.

FORMAL DESCRIPTION OF DECOMPOSITION TECHNIQUE

Definitions

Consider any rate (or more generally a ratio) where $R = P_1/P_2$.

Where P_1 = a collection of N_1 items representing set P_1

P_2 = a collection of N_2 items representing set P_2.

If there are i = 1 through K categories into which P_2 can be classified for which there is an associated category-specific rate, $r_i = P_{1i}/P_{2i}$, then R can be written as a weighted sum of the category specific rates, where each weight is the proportion of P_2 items in category i.

$$R = \sum_{i=1}^{K} \left(\frac{P_{2i}}{P_2}\right)\left(\frac{P_{1i}}{P_{2i}}\right) = \sum p_i r_i \qquad [1]$$

Equation [1] implies that a change in R, between time = x and time = x+y, can be written as

$$R_{x+y} - R_x = (\sum p_i r_i)_{x+y} - (\sum p_i r_i)_x \qquad [2]$$

Equation [2], in turn can be manipulated so that principal terms, representing the contribution of the components of change in R, can be isolated. This manipulation, which is the key to understanding this approach to decomposition analysis, can be illustrated with the following example.

$$\sum p_i r_i = (p_1 r_1 + p_2 r_2)$$

and a change in R could be written as

$$(\sum p_i r_i)_{x+y} - (\sum p_i r_i)_x = (p_1 r_1 + p_2 r_2)_{x+y} - (p_1 r_1 + p_2 r_2)_y \qquad [3]$$

which is algebraically equivalent to writing

$$(p_1 r_1 + p_2 r_2)_{x+y} - (p_1 r_1 + p_2 r_2)_x =$$

$$\left[\left[\frac{r_{1x+y} + r_{1x}}{2}\right](P_{1x+y} - P_{1x}) + \left[\frac{r_{2x+y} + r_{2x}}{2}\right](P_{2x+y} - P_{2x})\right] + \qquad [4]$$

$$\left[\left[\frac{P_{1x+y} + P_{1x}}{2}\right](r_{1x+y} - r_{1x}) + \left[\frac{P_{2x+y} + P_{2x}}{2}\right](r_{2x+y} - r_{2x})\right]$$

(See footnote 1 for a complete proof.)

Notice that there are two principal terms in the righthand side of Equation [4]. The first term gives the contribution of structural change in Factor P to the change in R while the second term gives the contribution of the category-specific rates, r_1 and r_2, to the change in R. Notice, further, that in the first term the "mean" values of r_1 and r_2 are taken over times x+y and x and multiplied by the differences (subtraction) for P_1 and P_2 between time x+y and x, respectively. In the second term, the "mean" values of P_1 and P_2 are multiplied by the difference in r_1 and r_2, respectively. This is the convention that gives the contribution of one factor--by taking its difference--while multiplying it against a "constant value" of the other factor--its mean value. Since this algebraic sleight-of-hand still sums to the original change in R, it provides a way to assess the contribution of each factor. This approach can, of course, be extended to include both more categories and more variables and in the formal definition given below, a three variable, multi-categorical decomposition is defined.

$$\Delta R = \sum_{i}^{k} \bar{r}_i \Delta p_i + \sum\sum_{ij}^{km} \bar{p}_i \bar{r}_{ij} \Delta p_{ij} + \sum\sum_{ij}^{km} \bar{p}_{ij} \bar{r}_{ij} \Delta r_{ij} \qquad [5]$$

Where $\bar{r}_i$, $\bar{P}_i$, $\bar{r}_{ij}$, and $\bar{p}_{ij}$ are averages and Δp_{ij}, Δp_{ij} and Δr_{ij} are differences, with p and r referring to a set of category-specific structural variables and category-specific ratios, respectively.

Equation [5] gives contributions to a difference (or change) in ratios from three effects, where each of its three terms refers to an effect. The first term gives the contribution of a difference in structural Factor I; the second term gives the contribution of differences in Factor J within Factor I; and the third term gives the contribution from changes in I-J specific ratios.

This three-term form will be used throughout the remainder of this paper.

SOME CHARACTERISTICS OF THE TECHNIQUE

This technique has, of course, some characteristics that make it less than ideal for component analysis. An important one is that the order in which variables are introduced affects the outcome. For example, if age is used as a primary structural factor and sex (by age) is a secondary structural factor, the effects of age and sex (by age) will be different than the effects of sex and age (by sex) for the same data, where sex is the primary structural factor and age (by sex) is the secondary. Another point is that the distinction between the contributions of structure and category-specific rates is not without ambiguity. Structural effects depend not only on differences but

also on the mean values of the category-specific rates, which, in turn, are determined by differences in these rates. Similarly, the effects of category-specific rates depend not only on differences between rates but also on the mean values of the structural effects, which, in turn, are determined by differences in these effects.

THREE EXAMPLE APPLICATIONS

Example 1: Changes in the overall state community college participation rate over time.

The first example is an age-sex decomposition of the change in the total community college participation rate (Fall headcount enrollment/total state population) for Washington between 1971 and 1979.

COMMUNITY COLLEGE
PARTICIPATION RATE (per 1000 persons)

Year	Rate
1979	49.71
1971	30.42

Change = 49.71 - 30.42 - 19.29

Components of Change (in percent)

Contribution from change in:

(1)	State population's sex-structure:	- 1.30	
(2)	State population's age-structure by sex:	13.70	
(3)	State participation rate structure:	88.13	
	Females		(69.52)
	Age 15-29		26.39
	Age 30+		42.46

Thus, we find that the net effect of changes in the state population's age-sex structure contributed to about 12 percent of the total change in the participation rate; while about 88 percent of the change can be attributed to increased participation. (That is, while holding constant the participation rate effects, demographic effects contributed about 12 percent; while holding

constant demographic effects, participation rate effects contributed about 88 percent.) Further, the increased participation rate of females, age 30 years and over, itself contributed to nearly 43 percent of the total change.

Example 2: Differences in participation rates between two community college service areas.

The second example uses Fall, 1979, participation rates from two areas within Washington state. In Eastern Washington, the community college participation rate was 57.38 (per 1000), while in the City of Seattle, it was only 54.91 (per 1000).

PERCENT DECOMPOSITION OF THE DIFFERENCE
IN TOTAL COMMUNITY COLLEGE PARTICIPATION RATE
BETWEEN EASTERN WASHINGTON AND SEATTLE

CONTRIBUTION FROM DIFFERENCE IN:

AGE STRUCTURE

All Ages	15-19	20-29	30+
26.50	0.10	42.70	-16.30

SEX-STRUCTURE BY AGE

Age	Both Sexes	Male	Female
15-19	-0.01	6.739	-6.748
20-29	0.130	-5.065	5.195
30+	-0.45	1.753	-2.203
All Ages	-0.33		

AGE-SEX PARTICIPATION RATES

Age	Both Sexes	Male	Female
15-19	187.70	80.30	107.40
20-29	-151.20	-108.20	- 43.00
30+	37.40	50.60	88.00
All Ages	73.90		

DATA UNDERLYING EXAMPLE 2

PROPORTION OF POPULATION BY AGE GROUP

Age	E. Washington	Seattle
15-19	.12859	.12852
20-29	.25990	.24926
30+	.61151	.62221

PROPORTION OF AGE GROUP THAT IS MALE

15-19	.52112	.49966
20-29	.51541	.52043
30+	.48409	.46290

PROPORTION OF AGE GROUP THAT IS FEMALE

15-19	.47888	.50034
20-29	.43459	.47957
30+	.51591	.53710

PARTICIPATION RATES (Both Sexes)

15-19	.07335	.04227
20-29	.09182	.10641
30+	.03820	.03689

PARTICIPATION RATES (Males)

15-19	.07545	.04522
20-29	.08775	.10802
30+	.03098	.03526

PARTICIPATION RATES (Females)

15-19	.08149	.03933
20-29	.09614	.10467
30+	.04498	.03829

This example shows that about one-fourth of the contribution to the higher participation rate found in Eastern Washington relative to Seattle is from differences in the age-structure: Eastern Washington has a higher proportion of people age 20-29, an age group with the highest participation rates. Within the contribution from participation rates, note that Seattle has higher rates for the age group 20-29, which provides strong negative effects. (That is, holding other factors constant, these effects contributed to a higher overall rate for Seattle.) However, the higher rates in Eastern Washington for age groups 15-19 and 30+ more than wiped out Seattle's advantage in the 20-29 age group.

Example 3: Changes in faculty/student ratios at a specific college over time

From 1978-79 to 1979-80, the ratio of faculty to students (FTE-F/FTE-S) declined from 18.46 to 17.23 at a Washington community college. This decline indicates that the college's staffing level in 1979-80 increased relative to its enrollment level.

We would like to know what has contributed to this decline. Accordingly, we will look at three major factors:

I. Shift in enrollment (FTEs) by time/location of courses

   A. Day-on-campus

   B. All other

II. Shift in enrollment (FTEs) by institutional intent of courses (within time/location categories)

   A. Academic

   B. Vocational

III. Shift in staffing by

   A. Day-on-campus

   B. All other

   C. Academic

   D. Vocational

These numbers show changes when controlling further effects of changes in other areas.

	ALL TIME/ LOCATION CATEGORIES	DAY-ON CAMPUS	NON-DAY-ON-CAMPUS
CONTRIBUTION FROM CHANGE IN:			
I. TIME/LOCATION FTES	+ 5.70	+18.14	-12.44
II. INSTITUTIONAL INTENT FTES			
A. Academic	+16.06	+54.66	+38.60
B. Vocational	+39.64	+70.73	-31.09
III. TIME/LOCATION - INTENT			
Specific FTE-F/FTE-S Ratio:			
A. Academic	+16.88	45.08	-28.20
B. Vocational	53.63	34.72	+18.91
TOTAL**	99.79	114.01	-14.22

**Total does not equal 100.00 because of rounding error.

DATA UNDERLYING EXAMPLE 3

	FTE-S	FTE-F	FTE-F / FTE-S
1978-79			
Day-On-Campus:			
Academic	289.17	21.20	.05447
Vocational	284.19	20.33	.07154
TOTAL	673.36	41.53	.06168
Non-Day-On-Campus:			
Academic	148.62	8.17	.05497
Vocational	362.17	14.45	.03990
TOTAL	510.79	22.62	.04428
GRAND TOTAL	1,184.15	64.15	.05417
1979-80			
Day-On-Campus:			
Academic	359.15	21.56	.06003
Vocational	339.99	26.07	.07668
TOTAL	699.14	47.63	.06813
Non-Day-On-Campus:			
Academic	182.34	8.59	.04711
Vocational	325.01	13.79	.04243
TOTAL	507.35	22.38	.04411
GRAND TOTAL	1,206.49	70.01	.05803

ANALYSIS OF EXAMPLE 3 RESULTS

In 1978-79, 56.86 percent of all FTE-students were comprised of day-on-campus FTE-students. In 1979-80, this increased to 57.95 percent.

In 1978-79, 57.80 percent of the day-on-campus FTE-students were academic; in 1979-80 this declined to 51.37 percent.

In 1978-79, only 29.10 percent of non-day-on-campus FTE-students were academic; by 1979-80, this share increased to 35.94 percent.

The day-on-campus increase between 1978-79 and 1979-80 was primarily due to increased vocational FTE-students, some of which also shifted from the evening and off-campus programs.

This decomposition of enrollment growth can then be helpful when looking at the overall ratio of faculty to students. When the strongest contributing factor to enrollment gorwth is an increase in the day-on-campus vocational program area, which by nature has fewer students per instructor, is accompanied by a decline in the academic area, it follows that the overall ratio of faculty to students will go up. Having a broader base of background information such as this provides for more interpretations than the quick conclusion that a higher faculty to student ratio implies a reduction in efficiency.

CLOSING COMMENTS

The examples cited are illustrative of how the decomposition technique can be used for analyzing differences over time between any two units--service areas, colleges, departments, and the like. In addition, the technique can be applied in an iterative fashion. For example, in the case of Faculty/Student ratios, the change in the ratio for day-on-campus academic programs can, itself, be decomposed by department, full-time/part-time instructor status, and so on.

BIBLIOGRAHPY

KITAGAWA, E.

1955 "Components of A Difference Between Two Rates." Journal of The American Statistical Association, 50 (December): 1168-1194

RETHERFORD, R. and L. J. CHO

1974 "Comparative Analysis of Recent Fertility Trends in East Asia." East-West Population Institute (Honolulu) Reprint No. 51.

FOOTNOTE 1

The easiest way to show this equivalency is to work backwards.

$$\left[\left[\frac{r_{1x+y} + r_{1x}}{2}\right](P_{1x+y} - P_{1x}) + \left[\frac{r_{2x+y} + r_{2y}}{2}\right](P_{2x+y} - P_{2x})\right]$$

$$+ \left[\left[\frac{P_{1x+y} + P_{1x}}{2}\right](r_{1x+y} - r_{1y}) + \left[\frac{P_{2x+y} + P_{2y}}{2}\right](r_{2x+y} - r_{2x})\right]$$

$$= (.5)[r_{1x+y} + r_{1x}](P_{1x+y} - P_{1x}) + (.5)]r_{2x+y} + r_{2x}](P2_{x+y} - P_{2x})$$
$$+ (.5)[p^{1x+y} + p^{1x}](r_{1x+y} - r_{1x}) + (.5)[P_{2x+y} + P_{2x}](r_{2x+y} - r_{2x})$$

Carry out some multiplication, we find that

$$(.5)[[(r_{1x+y})(P_{1x+y}) + (r_{1x})(P_{1x+y}) - (r_{1x+y})(P_{1x}) - (r_{1x})(P_{1x})]$$
$$+ [(r_{2x+y})(P_{2x+y}) + (r_{2x})(P_{2x+y}) - (r_{2x+y})(P_{2x}) - (r_{2x})(P_{2x})]$$
$$+ [(P_{1x+y})(r_{1x+y}) + (P_{1x})(r_{1x+y}) - (P_{1x+y})(r_{1x}) - (P_{1x})(r_{1x})]$$
$$+ [(P_{2x+y})(r_{2x+y}) + (P_{2x})(r_{2x+y}) - (P_{2x+y})(r_{2x}) - (P_{2x})(r_{2x})]]$$

Now, by rearranging the interior terms, some cancellations become apparent:

$$(.5)[(r_{1x+y})(P_{1x+y}) + (r_{1x+y})(P_{1x+y}) - (r_{1x})(P_{1x}) - (r_{2x})(P_{1x})$$

$$+ (r_{1x})(P_{1x+y}) - (P_{1x+y})(r_{1x}) - (r_{1x+y})(P_{1x}) + (P_{1x})(r_{1\,x+y})$$
$$+ (r_{2x+y})(P_{2x+y}) + (P_{2x+y})(r_{2x+y}) - (r_{2x})(P_{2x}) - (P_{2x})(r_{2x})$$

$$+ (r_{2x})(P_{2x+y}) - (P_{2x+y})(r_{2x}) - (r_{2x+y})(P_{2x}) + (P_{2x})(r_{2x+y})]$$

which, following the preceding cancellations, leaves

$$(.5)[(2)(r_{1x+y})(P_{1x+y}) + (2)(r_{2x+y})(P_{2x+y}) - (2)(r_{1x})(P_{1x}) - (2)(r_{2x})(P_{2x})]$$

which reduces to

$$[(r_{1x+y})(P_{1x+y}) + (r_{2x+y})(P_{2x+y})] - [(r_{1x})(P_{1x}) + (r_{2x})(P_{2x})]$$

which equals

$$(\xi p_i r_i)_{x+y} - (\xi p_i r_i)_x$$

AUG 27 1982

UNIVERSITY OF CALIFORNIA
ERIC® CLEARINGHOUSE FOR
JUNIOR COLLEGES
96 POWELL LIBRARY BUILDING
LOS ANGELES, CALIFORNIA 90024
EE 36